

加小于 2 dB/s) 的情况下^[3], 有良好的噪声估计能力, 能够通过递归平均法准确估计噪声值. 但该算法采用了固定搜索窗长来更新最小功率谱值, 当话音出现噪声谱陡增的情况时, 噪声估计会出现延迟. 延迟时间在 L 到 $2L$ 帧之间, 其中 L 为一个最小噪声估计窗长, 为了保证话音无失真, L 时间长度通常选为 800 ms~1500 ms.

为解决噪声估计延迟, 广大学者也做了一些研究. 2005 年, Loizou 提出在 MCRA 噪声估计算法的基础上引入连续平滑噪声更新^[4], 改进后的算法有效的缩短了噪声估计延迟, 但是该算法在低信噪比下存在噪声过估计情况, 导致话音失真. 2009 年, 贺慧芳, 马建芳在 MCRA 算法的基础上提出了一种基于语音帧检测的算法^[5]来加快噪声估计速度, 该算法在较高信噪比条件下可以有效的缩短噪声估计延迟, 在低信噪比下对噪声估计延迟的改进有限. 2015 年, 熊晶, 王旭贞等人在 MCRA 算法的基础上提出了一种双向噪声估计算法^[6,7], 该算法可以较好的解决噪声估计延迟以及过估计问题, 但是实现该算法需要在缓存一段语音数据, 不适合实时性要求较高的场合.

本文在 MCRA 噪声估计算法的基础上, 引入一种利用最大对数似然比结合能零比的 VAD 辅助算法, 通过该辅助算法, 可以动态的选择大小噪声搜索估计窗中的最小功率谱值进行后续计算, 从而得到一种变窗长搜索的改进型噪声估计算法. 本文所提的改进算法能有效缩短噪声估计延迟, 也能在低信噪比环境下运行, 并且不需要缓存数据, 可以用于实时性要求较高的场合.

2 基于变窗长的 MCRA 改进算法

本节先介绍 MCRA 算法^[2]的基本原理, 然后介绍本文提出的双参数双门限 VAD 算法以及双窗并行搜索, 最后讲述本文算法如何进行噪声估计更新.

2.1 MCRA 算法基本原理

当前带噪声语音帧的功率谱为 $|Y(k, l)|^2$, 式中 k 表示子带数, l 表示帧数, 后文中 k 和 l 含义同此处相同. 根据文献^[2], 先对当前帧功率谱做频域上的平滑:

$$S_f(k, l) = \sum_{i=-N}^N w(i) |Y(k, l)|^2 \quad (1)$$

其中, $w(i)$ 为一个长 $2N+1$ 的窗函数, 文中算法取 $N=1$.

然后对当前帧功率谱做时域上的平滑:

$$S(k, l) = \alpha_s S(k, l-1) + (1 - \alpha_s) S_f(k, l) \quad (2)$$

其中, α_s 为一个经验值, 文中算法 α_s 为 0.8.

再按照如下方式更新 $S_{\min}(k, l)$ 和 $S_{\text{tmp}}(k, l)$.

当 l 不能被 L (L 为定窗长) 整除时:

$$\begin{cases} S_{\min}(k, l) = \min(S(k, l), S_{\min}(k, l-1)) \\ S_{\text{tmp}}(k, l) = \min(S(k, l), S_{\text{tmp}}(k, l-1)) \end{cases} \quad (3)$$

当 l 能被 L 整除时:

$$\begin{cases} S_{\min}(k, l) = \min(S_{\text{tmp}}(k, l-1), S(k, l)) \\ S_{\text{tmp}}(k, l) = S(k, l) \end{cases} \quad (4)$$

$S_{\min}(k, l)$ 表示搜索窗内的最小功率谱值, $S_{\text{tmp}}(k, l)$ 用来缓存最小功率谱值.

接着求得当前帧平滑功率谱 $S(k, l)$ 及最小功率谱值 $S_{\min}(k, l)$ 的比值 $S_r(k, l)$:

$$S_r(k, l) = \frac{S(k, l)}{S_{\min}(k, l)} \quad (5)$$

如果 $S_r(k, l)$ 大于经验值 sigma (本文算法 $\text{sigma}=5$), 则带噪声语音子带语音存在标志 $I(k, l)$ 置 1, 否则 $I(k, l)$ 置 0, 并更新子带语音存在概率:

$$p(k, l) = \alpha_p p(k, l-1) + (1 - \alpha_p) I(k, l) \quad (6)$$

式中, α_p 为一个经验值, 本文算法中 α_p 为 0.9.

最后更新噪声估计值:

$$\lambda_d(k, l+1) = \lambda_d(k, l) p(k, l) + (1 - p(k, l)) (\alpha_d \lambda_d(k, l) + (1 - \alpha_d) |Y(k, l)|^2) \quad (7)$$

2.2 双参数双门限 VAD

(1) 基于能零比的 VAD

在文献^[8]中, Maragos 提出了一种基于短时能零比的 VAD 算法, 即短时平均过零率与短时能量相结合的 VAD 算法, 该算法对噪声谱变化不敏感, 即噪声谱的陡增不会影响算法性能, 且能较好的区分浊音和非浊音, 但在低信噪比下对清音和噪声的区分效果不佳.

短时平均过零率表示一帧语音中信号时域波形穿越横轴的次数. 对于离散的信号, 如果相邻的取样值发生符号改变, 则称为过零. 短时平均过零率就是样本数值改变符号的次数.

高频信号意味着较高的平均过零率, 低频意味着较低的平均过零率. 通常情况下, 语音中出现浊音时, 过零率较低, 而出现清音以及白噪声、超短波噪声时, 过零率较高.

过零率可以通过下面的公式求得:

$$Zcr(l) = \sum_{n=0}^{N-1} |\text{sgn}[y_l(n)] - \text{sgn}[y_l(n-1)]| \quad (8)$$

式中, l 代表帧数, n 表示数据点数, $\text{sgn}[]$ 为符号函数, $Z_{cr}(l)$ 为第 l 帧的过零率, $y_l(n)$ 为输入带噪声语音数据第 l 帧第 n 个数据点。

在语音中的有话区间能量要高于无话区间, 过零率则相反, 现在结合这两种方法, 即用能量值除以过零率的值, 可以更突出有话区间的数值, 而噪声区间的数值则更小. 第 l 帧的能量可以根据下面公式求得:

$$\text{Energy}(l) = \sum_{k=0}^{N-1} y(l, k) \quad (9)$$

这里引入对数能量计算公式:

$$\text{Log_Energy}(l) = \log_{10}(1 + \text{Energy}(l)/a) \quad (10)$$

用对数表达式可以缓和能量幅度剧烈变换, 参数 a 是一个经验值, 适当的调整好参数 a 的值, 将有利于噪声和清音的区分, 在本文中参数 a 设置为 2.

最后能零比通过以下公式求出:

$$\text{Ecr}(l) = \text{Energy}(l)/(Z_{cr}(l)+b) \quad (11)$$

参数 b 是为了防止 $Z_{cr}(l)$ 为 0 时出现异常值现象, 在本文中参数 b 设置为 1.

(2) 基于最大对数似然比的 VAD

在文献[9]中, J Chang 等人提出了基于高斯统计模型与拉普拉斯统计模型的最大对数似然比, 在文献[10]中, 夏丙寅等人基于上述模型提出了一种最大对数似然比 VAD 算法, 可以较好的区分清音和噪声. 为了保护清音部分, 本文选择最大对数似然比作为 VAD 辅助算法中的另外一种参数.

对每一帧语音信号, 语音存在假设为 H_1 , 语音不存在假设为 H_0 , 基于高斯统计模型, 可以求得带噪声语音信号傅里叶变换后每个频点所对应的似然度, 可以表示为:

$$\Lambda(k, l) = \frac{p(Y(k, l)|H_1)}{p(Y(k, l)|H_0)} = \frac{1}{\xi(k, l)} \exp\left\{\frac{\gamma(k, l)\xi(k, l)}{1 + \xi(k, l)}\right\} \quad (12)$$

$Y(k, l)$ 为第 l 帧带噪声语音信号的傅里叶变换的第 k 个子带处的频谱系数. $p(Y(k, l)|H_1)$ 是假设语音存在情况下的条件概率, $p(Y(k, l)|H_0)$ 是假设语音不存在情况下的条件概率. $\xi(k, l)$ 和 $\gamma(k, l)$ 分别为第 l 帧语音信号的第 k 个频点的先验性噪比和后验信噪比. 两种信噪比分别可由 DD 判决法则^[11]得到, 分别表示为以下公式:

$$\xi(k, l) = \frac{\lambda_s(k, l)}{\lambda_d(k, l)} \quad (13)$$

$$\gamma(k, l) = \frac{|Y(k, l)|^2}{\lambda_d(k, l)} \quad (14)$$

式 (6) 和式 (7) 中的 $\lambda_d(k, l)$ 表示噪声谱估计值.

基于对数似然比 LLR 的判决规则可得:

$$\text{LLR}(l) = \log \Lambda(l) = \frac{1}{N} \sum_{k=0}^{N-1} \log \Lambda(k, l) \quad (15)$$

本文中两种 VAD 算法结合起来, 利用能零比 VAD 对噪声功率谱变化不敏感, 以及最大对数似然比 VAD 能在低信噪比情况下较好的区分清音和噪声的特点, 提出一种双门限双参数的 VAD 算法, 并将其作为 MCRA 噪声估计算法的辅助算法.

上述双参数双门限的 VAD, 即以最大对数似然比以及能零比作为参数, 采用二级门限, 通过判决流程得到一个语音存在标记值. 具体的流程如图 1 所示.

状态 0 为“非语音段”、状态 1 为“疑似语音段”, 状态 2 为“语音段”, 状态 3 为“结束段”. 为方便描述, 此处用 $\text{Ecr}(l)$ 表示第 l 帧的短时能零比值, 用 $\text{LLR}(l)$ 表示第 l 帧的对数似然比值, $\text{Ecr}1$ 、 $\text{Ecr}2$ 分别表示二级能零比门限, $\text{LLR}1$ 、 $\text{LLR}2$ 分别表示二级对数似然比门限. $\text{LLR}1$ 取 3, $\text{LLR}2$ 取 0.4, $\text{Ecr}1$ 取值为前五帧无话段的平均能零比值的 1.3 倍, $\text{Ecr}2$ 取值为前五帧无话段的平均能零比值的 1.7 倍.

从状态 0 开始执行 VAD 算法. 如果当前状态为 0 或者 1, 且 $\text{Ecr}(l)$ 大于 $\text{Ecr}2$ 且 $\text{LLR}(l)$ 大于 $\text{LLR}1$, 就把该帧语音标记为状态 2, 并将无声帧计数器置 0, 语音帧计数器加 1. 若不满足该门限要求, 但 $\text{LLR}(l)$ 大于 $\text{LLR}2$, 则把该帧语音标记为状态 1, 语音帧计数器加 1. 如果上述两级判决都不满足则把该帧语音标记为状态 0.

如果当前状态为 2, $\text{Ecr}(l)$ 大于 $\text{Ecr}1$ 且 $\text{LLR}(l)$ 大于 $\text{LLR}2$, 则维持当前状态. 不满足上述要求, 无声帧计数器加 1. 若此时无声帧计数器的数值小于最长无声长度, 则认为当前帧依旧处于语音段, 语音帧计数器加 1. 若无声帧计数器的数值大于最长无声长度, 且语音帧计数器的值小于最小语音长度, 则将当前帧标记为状态 0, 语音帧计数器和无声帧计数器置 0, 否则将当前帧标记为状态 3.

最终根据语音帧所处的状态做一个存在标记, 当状态为 0 时, 把当前帧标记为“语音不存在”, 其他状态则把当前帧标记为“语音存在”.

2.3 双窗法并行搜索

原 MCRA 算法, 噪声搜索窗选用固定的窗长, 而

本文算法选用两个大小不同的窗^[12]同时搜索最小功率谱值,并根据式(3)和式(4)持续更新 $S_{\min}(k,l)$ 和 $S_{\min}'(k,l)$ 以及 $S_{\text{tmp}}(k,l)$ 和 $S_{\text{tmp}}'(k,l)$.在噪声估计更新中会根据上一节VAD算法所给出的语音存在标记来选择大窗或小窗下所搜索到的最小功率谱值,从而实现变窗长搜索.本文算法大窗 L 长度选用50帧、小窗 L 长度选用8帧.

2.4 噪声估计更新

若语音帧通过上述的双参数双门限VAD算法被标

记为“语音存在”,则根据式(5)计算当前帧平滑功率谱 $S(k,l)$ 与大窗下的最小功率谱值 $S_{\min}(k,l)$ 的比值 $S_r(k,l)$.若被标记为“语音不存在”,则根据式(5)计算当前帧平滑功率谱 $S(k,l)$ 与小窗下的最小功率谱值 $S_{\min}'(k,l)$ 的比值 $S_r(k,l)$.将 $S_r(k,l)$ 与经验值 σ 比较,确定 $I(k,l)$,最后根据式(6)和式(7)求得噪声估计值 $\lambda_d(k,l)$.

由于小窗长度较短,选用小窗下搜索到的最小功率谱值来进行噪声估计更新计算,会加快噪声估计速度.

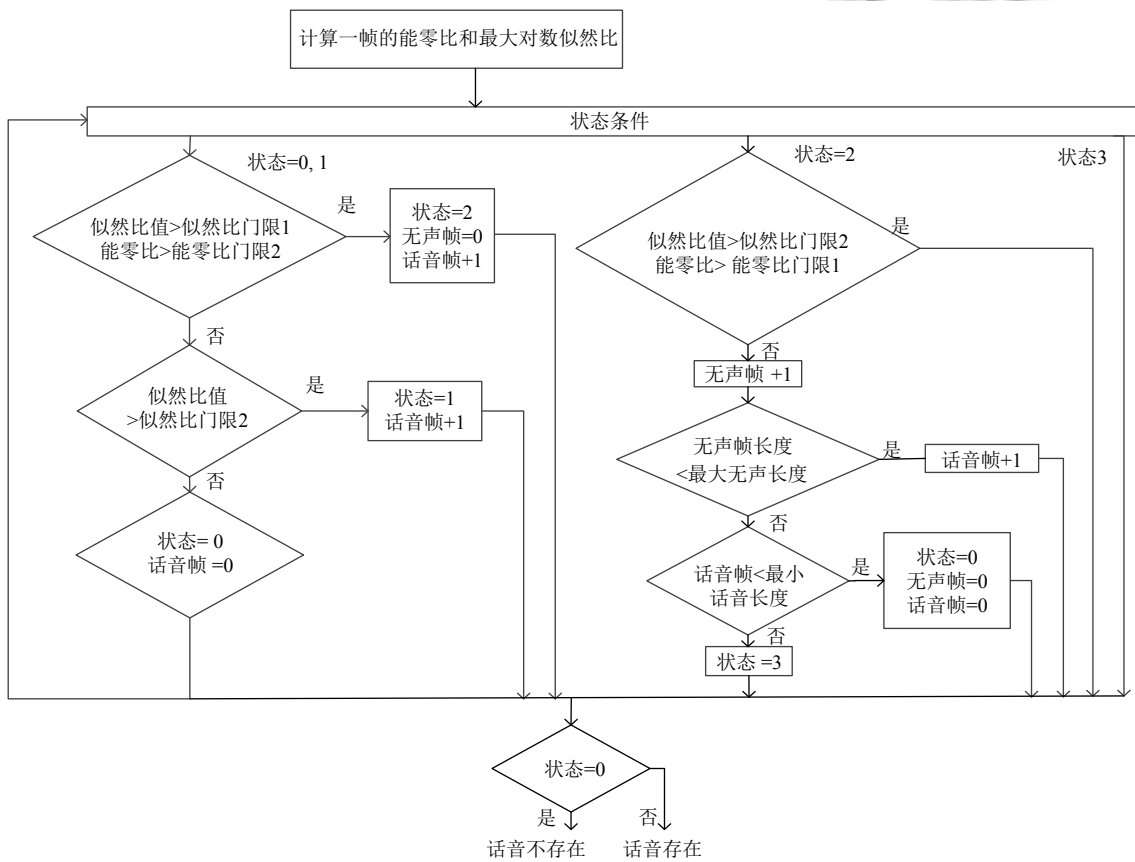


图1 辅助VAD算法流程框图

3 语音估计器

本文采用Israel Cohen提到的OM-LSA语音估计器^[13],结合本文的改进型噪声估计算法,实现整个语音增强算法,具体框图如图2所示.

图2中的 $y(n)$ 为带噪语音,加窗分帧后经过FFT,求得带噪语音的频谱 $Y(k,l)$,幅度谱 $|Y(k,l)|^2$,并通过本文算法求得噪声估计值 $\lambda_d(k,l)$.实验中分帧长度为32ms,重叠帧长度为16ms.

文献[13]中指出:通过估计每一个频率子带语音存在概率,得到一个联合时频特性的增益函数:

$$G(k,l) = \left(\frac{\xi(k,l)}{1+\xi(k,l)} \right)^{p(k,l)} G_{\min}^{1-p(k,l)} \quad (16)$$

其中 $\xi(k,l)$ 是第 l 帧第 k 个子频带的先验信噪比, $p(k,l)$ 是第 l 帧第 k 个子频带的语音存在概率,均根据噪声估计得到的噪声谱值来求解,具体求解方式参考文献[13],此处不做详细讨论.

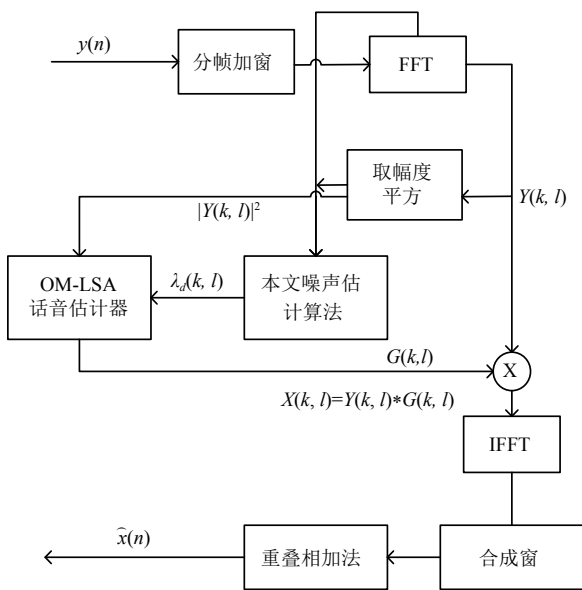


图2 语音增强框图

通过下式求得处理后的语音频谱:

$$\hat{X}(k, l) = Y(k, l) * G(k, l) \quad (17)$$

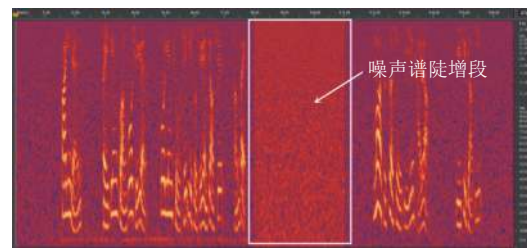
其中, $\hat{X}(k, l)$ 为处理后的语音频谱, 最后通过 IFFT 快速傅里叶反变换得到时域表达式 $\hat{x}(n)$, 合成输出得到最终处理结果.

4 实验仿真

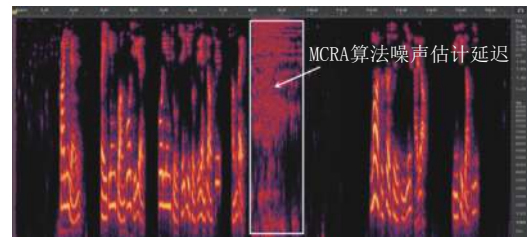
本节的实验仿真内容主要为: 将引入双门限双参数 VAD 辅助算法的变窗长改进算法与原 MCRA 算法^[2] 做对比, 比较收敛时间长短. 实验中的噪声类型采用选自 Noisex 中的高斯白噪声, 以及超短波噪声, 然后分别将高斯白噪声、超短波噪声与纯净语音叠加, 生成几段样本用来测试. 实验选取标准普通话测试语音 17 s, 采样率 8 KHz.

4.1 实验设计

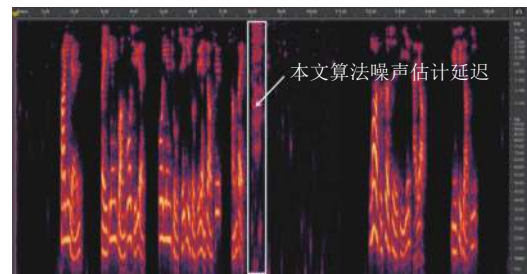
实验一. 选取上述纯净语音样本, 给样本加上高斯白噪声, 带噪语音 8 s-11 s 处信噪比由 10 dB 变为 2 dB, 在语谱图 3(a) 中用方框标出, 并用文字说明. 图 3(b) 为原 MCRA 算法在噪声谱陡增时的噪声估计延迟, 图中用方框标示, 可以看到噪声估计延迟较长, 约为 1.6 s; 而图 3(c) 为本文改进算法的语谱图, 可看到噪声谱陡增处, 噪声估计延迟缩短, 约为 0.4 s.



(a) 带噪语音



(b) 原MCRA算法噪声估计延迟



(c) 本文改进算法噪声估计延迟

图3 高斯白噪声环境下的算法对比

实验二. 选取上述纯净语音样本, 给样本加上超短波噪声, 此环境下噪声谱陡增段包含语音, 在 2.5 s-4 s 处, 信噪比由 10 dB 变为 0 dB, 语谱图 4(a) 中用方框标出, 并用文字说明. 图 4(b) 中可以看到, 噪声估计延迟约为 1.5 s; 而图 4(c) 为文中改进算法的语谱图, 可以看到噪声谱陡增处, 噪声估计延迟缩短, 约为 0.5 s.

4.2 实验结果分析

从图 3 和图 4 中, 可以看出: 无论是在高斯白噪声还是超短波噪声环境下, 本文提出的改进型算法, 面对噪声谱陡增, 噪声估计延迟都要比原 MCRA 算法短, 即噪声估计速度本文提出的算法要优于原 MCRA 算法.

5 总结

本文利用了过零率对噪声功率谱变化不敏感的性质, 将能零比与最大对数似然比结合, 作 VAD 辅助检测. 联合原有算法, 采用大小窗同时搜索功率谱最小值, 最终形成基于变窗长搜索的改进型噪声估计算法. 经过实验证明, 改进的噪声估计算法, 在噪声谱陡增时,

能够有效的降低噪声估计延迟。

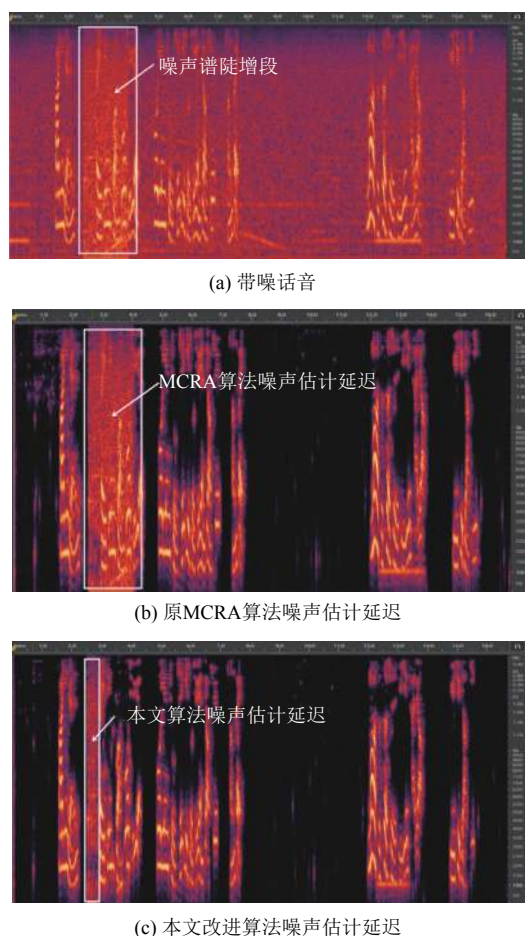


图4 超短波噪声环境下的算法对比

参考文献

- Martin R. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on Speech and Audio Processing*, 2001, 9(5): 504–512. [doi: 10.1109/89.928915]
- Cohen I, Berdugo B. Noise estimation by minima controlled recursive averaging for robust speech enhancement. *IEEE Signal Processing Letters*, 2002, 9(1): 12–15. [doi: 10.1109/97.988717]
- 郑成诗, 李晓东, 莫福源, 等. 一种噪声谱估计的快速跟踪算法. *中国声学学会 2006 年全国声学学术会议论文集*. 厦门, 中国. 2006. 401–402.
- Rangachari S, Loizou PC. A noise-estimation algorithm for highly non-stationary environments. *Speech Communication*, 2006, 48(2): 220–231. [doi: 10.1016/j.specom.2005.08.005]
- 贺慧芳, 马建芬. 基于语音帧检测和子带谱跟踪的噪声估计算法. *计算机工程*, 2009, 35(20): 281–282. [doi: 10.3969/j.issn.1000-3428.2009.20.099]
- 熊晶. 语音增强中噪声估计的研究[硕士学位论文]. 兰州: 兰州交通大学, 2015.
- 王旭贞. 汽车噪声下语音增强算法的研究与应用[硕士学位论文]. 杭州: 杭州电子科技大学, 2015.
- Maragos P. Fractal aspects of speech signals: Dimension and interpolation. *Proceedings of 1991 International Conference on Acoustics, Speech, and Signal Processing*. Toronto, ON, Canada. 1991. 417–420. [doi: 10.1109/ICASSP.1991.150365]
- Chang JH, Kim NS, Mitra SK. Voice activity detection based on multiple statistical models. *IEEE Transactions on Signal Processing*, 2006, 54(6): 1965–1976. [doi: 10.1109/TSP.2006.874403]
- 夏丙寅, 鲍长春. 适应噪声强度突变的噪声估计加速方法. *信号处理*, 2013, 29(10): 1336–1345. [doi: 10.3969/j.issn.1003-0530.2013.10.010]
- Ephraim Y, Malah D. Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1984, 32(6): 1109–1121. [doi: 10.1109/TASSP.1984.1164453]
- 张建伟, 陶亮, 周健, 等. 基于改进谱平滑策略的IMCRA算法及其语音增强. *计算机工程与应用*, 2017, 53(1): 153–157. [doi: 10.3778/j.issn.1002-8331.1504-0029]
- Cohen I, Berdugo B. Speech enhancement for non-stationary noise environments. *Signal Processing*, 2001, 81(11): 2403–2418. [doi: 10.1016/S0165-1684(01)00128-1]